

Задание 2

Корреляционный анализ, и равенство средних

В результате проведения эксперимента получены данные (файл data1.csv).
Считать, что величина var5 зависима от остальных (var0-var4).

Для набора данных 1 (data1.csv):

1. Построить матрицу корреляции.
2. Проверить гипотезы о значимости каждого коэффициента каждой величины с пятой (var5).
3. Построить диаграмму рассеивания для var5 и двух других величин: одну для самой слабой корреляции, другую для наиболее сильной.

В эксперименте измерялись значения двух величин. Были получены зависимые выборки

Для этого набора данных 2 (data2.csv):

1. Построить диаграммы размаха для выборок.
2. Найти 95% доверительный интервал для средних значений выборок, полагая что генеральные совокупности имеют нормальное распределение.
3. Проверить гипотезу о равенстве средних значений для двух выборок.
4. Проверить гипотезу о равенстве двух выборок используя непараметрический критерий

Вопросы

1. Что такое выборка? Что такое генеральная совокупность?
2. Что такое корреляция?
3. Что такое коэффициент корреляции (линейный)?
4. Что такое статистическая гипотеза?
5. Что такое p-value? Как это значение используется для проверки гипотез?
6. Почему нужно проверять гипотезу о равенстве коэффициента корреляции генеральной совокупности нулю?
7. Что такое диаграмма размаха?
8. Что такое матрица корреляции?